1 Data Processing And Information

1.1 Data and information

- **data**: It is raw or unprocessed data which has no meaning
- **information**: It is data with context and meaning

Direct and Indirect Data

- direct data is data that has been collected from the source for a specific purpose and used for only that purpose
- indirect data is data that is obtained from a third party and used for a different purpose than that which it is was originally collected for

Sources of Direct Data:

- **questionnaires:** They are a set of arranged questions that can be distributed to people online or physically in order to collect data from individuals. They are user friendly, and easier to distribute and analyse information since all respondents answer same questions.
- **interviews:** They are one on one meeting between interviewee and interviewer, in which questions are asked by an interviewer to the interviewee to get information, they can be close ended questions which can make it easier to quantify the responses or they can be open ended questions to get more in depth detail for better quality data.
- **observation:** A method of collecting data by watching an event/activity and then analyzing it and recording the data. The observer gets the information firsthand rather than from a thirdparty.
- **data logging:** The use of sensors and computers to gather and anaylse data allowing it to be saved or represented as an output in graphs,charts, etc. Its most likely to be used in scientific experiments where human intervention isn't suitable such as a situation which requires continuous monitoring and gathering of data

Sources of Indirect Data:

- **electoral registers:** It's a record of citizens that are eligible to vote in an election, it contains a list of their personal data such as legal name, address, contact information, etc. Some of the is data removed from open version of the register which is accessible to organizations and public for certain use depending on the local laws.

- **businesses collecting info when used by third parties:** Businesses sell the information that they collect from their customers. For example when someone purchases something online they are often asked to tick a box authorising the business to share this with other organisations. Customers often provide personal information that has a commercial value. Businesses use this information to create mailing lists that can be purchased by any other organisation/individual to send emails or even brochures through the post.
  Note: business collect the data from customers as direct data for shipping their products/providing services etc, this data can be purchased by third parties hence considered as indirect data now, it can be used to for targeting required audience (customers) / analyzing buying trends etc.

Advantages of Direct Data:

- Source of data is known hence we know how reliable it is
- Only required data is gathered hence its very relevant
- Data can be sold later as indirect data for other purposes
- Data is likely to be upto data
- Data can be easily presented in required format as preferred and suitable source is used

Disadvantages of Direct Data:

- It can take a long time to gather data than to acquire data from already existing indirect data source
- Larger samples can be difficult to collect.
- Can be more expensive than indirect data due to the preparation and gathering of required data; such as producing questionnaires or buying additional equipment such data loggers
- Data might be out of data when the project is completed

Advantages of Indirect Data:

- The data is readily available
- Allows larger set of data to be examined with less time and cost involved in comparison with direct data
- A larger sample size can be used
- Allows data to be gathered from subjects (eg; people) which the gatherer doesn't have physical access to

Disadvantages of Indirect Data:

- Can be less reliable as source maybe unknown
- Not all data will be relevant
- Might be out of date
- Might be difficult to extract data as it would be in wrong format
- There may be sampling bias (data collected certain purpose from only certain sample would not give an accurate answer for the purpose its being used now)

1.2 Quality Of Information

Factors affecting the quality of information are:-

- **accuracy:**
  If the data collected is inaccurate, the information after processing will be inaccurate and hence of bad quality. Misspelling words or misplacing characters could lead to inaccuracy, i.e 10:30 am for 10 o clock at night
- **relevance:**
  Data must be relevant to the purpose, irrelevant data needs to be removed before processing for better quality information, i.e; being given a bus timetable when train timetable is required
- **age:**
  Information must be upto date, old information will be irrelevant and inaccurate and hence of bad quality, i.e, not updating family registers will make the emergency contacts incase of emergency useless due to outdated information
- **level of detail:**
  Information must be of required level of detail. Too much detail will make it difficult to extract necessary information and too little detail will not provide the information needed

- **completeness:**
  Information must be complete and have all required information to be of good quality, if not it can't be used properly for a particular purpose, i.e not having the venue of an event mentioned in its advert poster makes it incomplete. (note: information can have higher level of detail and be complete)

1.3 Encryption

- **encryption:** It's the process of converting plain text into cipher text which makes the original data unintelligible

The need for encryption:

- Encryption is important when sending or storing sensitive data such as personal data or a company's sale figure
- Data being sent across a network or the internet can be easily intercepted by hackers
- Data stored on storage media could be stolen or lost
- Hence the purpose of encryption is to scramble the data in order to make it difficult or impossible to read if it is accessed by an unauthorized user

Methods Of Encryption:-

- **Symmetric:**
  A method of encryption which requires the use of the same private key in order to encrypt and decrypt data. The sender and receiver both require the same key, hence it needs to be agreed on before transmission of data or sent along with the files.
- **Asymmetric:**
  A method on encryption that requires the use of a public key (available to anyone) to encrypt data and private key (known only to recipient) to decrypt data. The same key can't be used to decrypt if it is used for encryption and vice versa.

Encryption Protocols:

An encryption protocol is a set of rules setting out how the algorithms should be used to secure information. There are several protocols including:

- IPsec (internet protocol security)
- SSH (secure shell)
- Transport Layer Security (TLS) and Secure Socket Layer (SSL)

TLS and SSL:

It is the most popular protocol used when accessing web pages securely. TLS is an improved version of SSL and has now, more or less, taken over from it.

Three main purpose of SSL/TLS:

- Enable encryption in order to protect data
- Make sure that the people/companies exchanging data are who they say they are (authentication)
- Ensure the integrity of the data to make sure it has not been corrupted or altered

The use of SSL/TLS in client server communication:
TLS is used for applications that require data to be securely exchanged over a client – server network, such as web browsing sessions and file transfers. In order to establish a connection between the client and server, a handshake needs to take place which authenticates the server to the client before the transfer of data can take place. It allows both parties to agree on a set of rules for communication and authenticate each other aswell as communicate securely through asymmetric and symmetric encryption.

Uses of encryption:

- **hard disk encryption:**
  - When a file is created/written on a disk, it is automatically encrypted, when it is read, it is automatically decrypted while leaving other files encrypted.
  - The whole disk is encrypted so that data is protected if the disk is stolen or left unattended.
  - Keys need to be secured in a well available location as data can't be recovered without the key.
  - Data can be permanently lost if the encrypted disk crashes or the OS gets corrupted.
- **email encryption:**
  - Email encryption uses asymmetric encryption, hence both the sender and recipient needs to send each other a digitally signed message to add each other's digital certificate to contacts.
  - Encrypting emails also encrypts any attachments.
  - Emails are susceptible to being intercepted by hackers, therefore encrypting all emails including downloaded ones are a good practice.
- **encryption in https websites:**
  - Hyper Text Transfer Protocol is shown by a URL having https:// or a padlock.
  - A session key is encrypted using a public key which is sent to the webserver by web browser, then it is decrypted using the servers private key, after which all exchange of information is conducted through encryption using a the session key.
  - HTTPS uses asymmetric encryption initially to establish a secure session, then uses symmetric encryption after that point forward
  - After the session is ended, the symmetric key is disposed of.

- Slower than http and needs to be kept upto date by the host, but is more secure in data transfer and sites with https are give more priority by search engines.

1.4 Checking The Accuracy Of Data

- **validation:** The process of checking data to make sure it matches the acceptable rules
- **presence check:** Checks if data is present and entered
- **range check:** Ensures if data is within a defined range, it has an upper and lower boundary
- **type check:** Ensures that the data is of a defined type
- **length check:** Ensures if data is of specified length
- **format check:** Ensures if data is of specified format such as data in dd/mm/yyyy
- **check digit:** Uses an algorithm to create a digit or character from the given data and attaches it to it as the last digit, this is then recalculated to check if the initial digits are entered correctly
- **look up check:** Checks if the data entered is within the list
- **consistency check:** It checks if the data is consistent with the other selected fields, such as password confirmation in signup forms
- **limit check:** Checks if the data is within the specified range but it only has one boundary, such as 13+ age ride check at parks
- **verification:** It is the process of checking that the entered data matches the original source
- **visual checking:** A manual check performed by the user entering the data. After data entry is complete, the data on screen is compared against the original document/source and any errors are corrected before proceeding (i.e banking apps confirming from the user the amount entered for money transfer)
- **double data entry:** A method of entering the data twice and then comparing the two entries, if they don't match then an error has occurred. (the verification could be performed by another person or done by a computer, i.e entering password twice during sign up)
- **parity check:** A method of verification to check whether the data has been changed or corrupted following a data transmission from one medium to another. A byte for example is assigned a parity bit, the parity is decided at start and can be either odd or even. If it's an odd parity then parity bit is either assigned 1 or 0 to make the total numbers of 1s in the parity byte odd. The same is done if even parity is used
- **checksum:** A method to check if data has been changed or corrupted following data transmission. The data is sent in blocks and an additional value called the checksum is calculated by using algorithm such as a hash function on the data. The checksum is send

along at the end of the blocks of data and recalculated again at the receivers end. Both of the checksum values are compared and if they don't match an error had occurred during the transmission

- **hash total:** A hash total is calculated before the transmission of data by adding up all numbers in one or more selected fields. The hash total is sent along data and is recalculated at receivers end, if the hash total are same then data has been transmitted correctly. If the selected field is alphanumeric, it is converted to number for the purpose of hash total and then added up together.
- **control total:** It is similar to hash total and performs the same action but the calculation is done on numeric fields only, the value produced is meaningful as alphanumeric fields are not used. (i.e if marks field is used as control total, then this control total value can be meaningful and used to find a class average marks for that test)

The need for both verification and validation:

- Validation is always carried out by a computer whereas verification can be carried out by a human or a computer
- Validation is checking that the data entered is reasonable and sensible
- Verification is checking that the data has been entered, copied or transmitted correctly but it doesn't tell if its sensible or not
- (add example how with verification and validation a data can be still incorrect if the original data is incorrect or swapped but in correct format etc)
- Verification is a way of ensuring that the user doesn't make a mistake when inputting data whereas validation is checking that the data input conforms with what the system considers to be sensible and reasonable
- By using both the chances of entry data incorrect can be reduced

1.5 Data Processing

Data processing is when data is collected and translated into usable information. Data processing starts with data in its raw form and translates it into a more readable format such as graphs, diagrams and reports. The processing is required to give data structure and context necessary so it can be understood by other companies and then used by employees throughout an organization. Data processing includes actions such as:

- collection and storage
- editing and updating
- sorting and searching

- output and dissemination

Batch Processing

In a batch processing system, the individual operations or transactions that need to be performed on the data are collected together into a batch and then processed at a later date instead of being worked on one by one by an operator in real time. The data is searched using sequential access

Examples:

- automated backups
- the processing of employees wages
- customer orders
- stock control

It makes use of 2 files:

- **master file:** It stores important data that doesn't change often such as a person name, number and address and is sorted in order of keyfield
- **transaction file:** It stores data that requires frequent changes that could be weekly or daily changes such as hours worked, items sold today, number of visitors

In order to update the master file, a new blank file will be created and used as the new master file. The following basic algorithm is used.

ID	Transaction	Employee name	Department
2	D	Julia Bolero	Sales
4	C	Nigel Ndlovu	Buying
7	D	Adrienne Pascal	IT
11	A	John Ward	Stores
12	A	Paolo Miserere	IT
EOF	End of file marker		

ID	Employee name	Department
1	Jose Fernandez	Buying
2	Julia Bolero	Sales
3	Louis Cordoba	Sales
4	Nigel Ndlovu	Stores
5	Bertrand Couture	Buying
6	Lionel Sucio	Stores
7	Adrienne Pascal	IT
8	Gurjit Mandare	Stores
9	Iqbal Sadiq	IT
10	Tyler Lewis	Buying
EOF	End of file marker	

```
First record in the transaction file is read
First record in the old master file is read
REPEAT
    IDs are compared
    IF IDs do not match, old master file record is
    written to new master file
    IF IDs match transaction is carried out
        IF transaction is D, old master file record
        is not written to new master file
        IF transaction is C, data in transaction file
        is written to new master file
    IF IDs match, next record from transaction file
    is read
    Next record from master file is read
UNTIL end of old master file
Data in transaction file record is written to new
master file
Any remaining records of the transaction file are
written to the master file
```

ID	Employee name	Department
1	Jose Fernandez	Buying
3	Louis Cordoba	Sales
4	Nigel Ndlovu	Buying
5	Bertrand Couture	Buying
6	Lionel Sucio	Stores
8	Gurjit Mandare	Stores
9	Iqbal Sadiq	IT
10	Tyler Lewis	Buying
11	John Ward	Stores
12	Paolo Miserere	IT
EOF	**End of file marker**	

Use of batch processing in payroll:

ID	Department	Rate ($)	Wages_to_date
036	Sales	20	1280
047	Buying	25	1475
165	Buying	25	1525
469	Sales	20	1160
512	Stores	15	825
545	Sales	20	1220
578	IT	30	1860
682	Sales	20	1080
778	IT	30	1920
786	Buying	25	1575
789	IT	30	1830
861	Stores	15	795
EOF	**End of file**		

ID	Hours_worked
036	40
469	40
578	38
778	40
789	40
EOF	**End of file**

```
1  First record in the transaction file is read
2  First record in the old master file is read
3  REPEAT
4     IDs are compared
5     IF IDs do not match, old master file record is
      written to new master file
6     IF IDs match transaction/calculation is carried out
7        Computer calculates the pay, Rate (from master file)
         multiplied by hours worked (from transaction file)
8        Wages _ to _ date is updated and record is written
         to new master file
9     IF IDs match, next record from transaction file
      is read
10    Next record from master file is read
11 UNTIL end of transaction file
12 Remaining records of the master file are written to
   the new master file
```

Use of batch processing with customer orders:

Cust_no	New_orders	Payment_made
219	320	200
451	870	1500
523	190	340
834	520	250

▲ Table 1.11 Transaction file

Cust_no	Balance
138	0
187	0
219	0
451	*-800*
487	*-260*
523	*-340*
764	0
802	*-920*
834	0
869	*-540*

▲ Table 1.12 Master file

```
1  First record in the transaction file is read
2  First record in the old master file is read
3  REPEAT
4     Cust _ nos are compared
5     IF Cust _ nos do not match, old master file record
      is written to new master file
6     IF Cust _ nos match transaction/calculation is
      carried out
7        Computer calculates New _ orders minus Payment _
         made and subtracts from Balance
8        Balance is updated and record is written to new
         master file
9     IF Cust _ nos match, next record from transaction
      file is read
10    Next record from master file is read
11 UNTIL end of transaction file
12 Remaining records of the master file are written to
   the new master file
```

Advantages of Batch Processing:

- It is a single, automated process requiring little human participation which can reduce costs
- Processing can be scheduled when there is little demand for computer resources, for example, at night, allows to get more work out of hardware
- As it is an automated process, there will be none transcription and update errors that human operators would produce
- There are fewer repetitive tasks for the human operator

Disadvantages of Batch processing:

- Only data of same type can be processed since an identical, automated process is being applied to all the data
- Errors cannot be corrected until the batch process is complete
- Information is not upto date unless until the master file has been updated by the transaction file

Online Processing

Online processing is done in a computer that has direct communication with a user. Data is processed almost immediately with a short delay and output is provided instantly, making it seem like the user is in direct communication with the computer. Each transition is processed before the next transaction is dealt with. Data is searched using direct access.

Electronic Funds Transfer (EFT):

Process if sending money from one bank account to another using computer software and without the involvement of banks staff, eg; ATM, online banking.

```
1  Select transfer money
2  Select the account you wish to transfer money from
3  Select new payee
4  Type in sort code, account number and payee name
5  Type in amount to transfer
6  Computer checks available balance
7  If you have sufficient funds, the transaction is
   authorised
8  Your bank's computer contacts payee's bank's
   computer which searches for the payee's record
9  Amount is subtracted from your account balance
10 Amount is added to payee's account balance
```

Electronic Funds Transfer at Point of Sale (EFTPOS):

Customer going to a point of sale, i.e; going to the counter for checking out / waiter bringing the card machine to table for payment, is considered a Point Of Sale.

1. Card chip is read and checked to make sure it is in date and it is a valid card number
2. If not, card is rejected and transaction terminated
3. PIN is entered by customer into PIN pad
4. Chip reader determines PIN from the chip
5. The two PINs are compared
6. If they are identical the transaction is authorised
7. If they are not identical, error message appears on the chip reader and two more attempts are allowed
8. If the two PINs are still not identical, the transaction is rejected and error message issued
9. Customer's bank is contacted by supermarket's computer
10. Customer's bank retrieves customer's record
11. Customer's bank checks if sufficient funds in account
12. If there are insufficient funds then transaction is rejected
13. If there are sufficient funds then transaction is authorised
14. The amount of the bill is deducted from customer's account
15. The amount of the bill is credited to supermarket's account

Automatic Stock Control:

An automated system which manages stock control with little human input.

1. The product's barcode is input from the barcode reader
2. The computer searches for this barcode number in the product file and finds it using direct access
3. The number in stock is reduced by one
4. The computer then compares the number in stock with the re-order level
5. If the number in stock is not equal to the re-order level then go back to step 1 and repeat
6. If the number in stock is equal to the re-order level then the computer creates an automatic order
    7. It looks up the re-order quantity of that product
    8. It looks up the supplier number of that product
    9. It searches the supplier file for the record corresponding to the supplier number found in the product file
    10. It sends the order automatically to the supplier using the supplier's contact details
11. Go back to step 1 and repeat

Electronic Data Exchange:

Electronic data exchange or electronic data interchange (EDI) is a method of exchanging data and documents without the use of paper. The documents can take any form such as invoice or order with the electronic exchange through computers using a standard format.

An EDI generally has these steps:

1. A company decides to buy some goods, creates an order and does not print it
2. EDI software creates an electronic version of the order and sends it automatically to the supplier
3. Supplier's computer system receives the order and updates its system
4. Supplier's computer system automatically sends a message back to the company, confirming receipt of the order

Business To Business Buying and Selling:

- Refers to buying and selling between two businesses.
- A B2B marketplace is similar to B2C marketplace in terms of appearance
- On B2B marketplaces, bulk orders can be placed and they can be edited online
- Buyers can compare products from different sellers, receive testers/samples and receiver discounts on large orders
- Sellers can save time and money that would have used to setup a large website, on marketing their products and they can do mini test sale runs to see if a product sells well
- B2B marketplaces has more government regulations and complex taxation, shipping is also complicated and expensive for large orders

Online Stores:

- Online stores are websites for a certain shop/chain to sell their products and services online.
- Orders are placed by like how its done in real life by browsing the online catalogue and adding selected items in a virtual cart and hence checking out online.
- Customers can look at a wide range of shops online and compare prices
- Customers don't need to spend extra money on travelling making the shopping online cheaper and faster
- Items are usually cheaper since no on street store is required and wages for staff is cheaper
- Shopping can be done at convenience without being rushed

- Reviews for services and products can be found instantly online

Method for checking out online:

```
1  IF user is new customer, they must register
2     Enter a username
3     Enter and verify (by double entry) password
4     Enter phone number and email address
5     Enter delivery/shipping address
6     Enter billing address if different to shipping address
                                                    [some sites only]
7     Choose speed/cost of delivery                 [some sites only]
8     Enter type of card and credit/debit card number
9     Enter date of expiry
10 IF user is existing customer
11    Log on by entering username and password
12    Confirm delivery address
13    Choose speed/cost of delivery                 [some sites only]
14    Select credit card/debit card account to be debited if stored,
      otherwise enter type of card and credit/debit card number
15 Enter card security code              [some sites may not require this]
16 Confirm order
```

Advantages of Online System:

- Easier to maintain and upgrade as banks, etc have less busy times so it can be shutdown for maintenance
- Errors are revealed immediately allowing it to be worked on immediately
- Useful for online money transactions
- Useful in online shopping
- Support and stability

Disadvantages of Online System:

- Lots of online requests can be difficult to manage as some are spam which can cause system to crash
- May require specialized staff to manage the online systems which increases costs
- Failure of network can cause the system to go down
- Requires entry of information immediately, making it expensive to run the system

Real Time Processing

Real time processing system is where data is processed as soon as it has been processed and output is generated immediately. The processing takes places continuously and only stops when system is turned off by user.

Examples:

- computer games
- traffic lights
- green houses

Some real time systems use a feedback loop where the output directly affects the input. It makes use of a microprocessor and sensors, sensors measure physical variables and send it to the microprocessors which compare it with a stored value. If its greater than stored value then microprocessors sends control signals to an actuator which turns off/on the [any device] . This immediatly affects the new readings sensors picks up. i.e air conditioning systems. Feedback is basically when the output of the system affects the new input, increasing ac temp will increase the temp of room and hence the new inputs will differ.

Air- Conditioning Systems:

1. User enters required temperature using keypad/ touchscreen
2. Microprocessor stores required temperature as a pre-set value
3. Microprocessor receives temperature from sensor
4. Microprocessor compares temperature from sensor to pre-set value
5. If temperature from the sensor is higher than the pre-set value, the microprocessor sends a signal to an actuator to switch the fans on
6. If temperature is lower than or equal to the pre-set value the microprocessor sends a signal to an actuator to switch the fans off
7. This sequence is repeated until the system is switched off

Rocket Guidance System:

A rocket guidance system makes use of real time processing. As the rocket is launched it could veer off course (divert from path) and hence crash. This is where the sensors come in and measure the respective variable and send it back to the microprocessor which compares it against stores values. The microprocessor sends appropriate control commands to actuator immediately to rotate the rocket back to course. Here the output (rotating rocket) affects the new input (rocket back at path) to the control system. As the rocket moves, its position also constantly changes so the processing is done continuously to ensure rockets stays in path or readjusts paths according to the situation, any delay in receiving instruction can cause the rocket to veer off or crash. This guidance system hence provides stability for the rocket and controls its movement.

Advantages of Real Time Processing:

- has fast real time analysis/processing
- information is always up to date, allows computer/microprocessor to take immediate action
- data is collected instantaneously

Disadvantages of Real Time Processing:

- occupies the CPU constantly, hence it can be expensive (uses constant power)
- requires expensive and complex computer systems
- difficult to maintain as it has no down time